USES OF DATA FROM THE SURVEY OF INCOME AND EDUCATION FOR POLICY ANALYSIS

Alan Ginsburg and George Grob
Office of the Assistant Secretary for Planning and Evaluation, DHEW

Introduction

The Federal government uses data in two ways -- for program administration and for policy analysis. Administrative uses of data are generally sanctioned by law or regulation, and involve the allocation of program resources under established formulas. Such uses are essentially routine and afford little or no room for the exercise of discretion; program entitlements are automatically determined "by the numbers." In contrast, policy analysis, which involves the uses of data to define and evaluate alternative courses of action, is much more episodic and judgmental in character.

Occasionally, when policy analysts focus on existing allocation formulas, the two types of use intersect. Under these conditions, data are used analytically to evaluate other administrative uses of data. Such has been the case with the Survey of Income and Education (SIE). The survey was expressly authorized by Congress with a view to evaluating the continued use of 1970 Census poverty statistics in the allocation of funds under Title I of the Elementary and Secondary Education Act.1/

The question of the likely impact of the SIE data on Title I allocations will be examined in the second half of this paper. Here, we wish to consider briefly the variety of Federal programs already using data similar to that now available from the SIE, and to describe some of the specific ways in which the SIE data lend themselves to policy studies.

Administrative Uses

Overall, programs in at least five departments -- Agriculture, HEW, HUD, Labor, and Treasury -- utilize income or employment statistics in current allocation formulas, and thus, are potential users of the SIE data. In terms of total funding, the Revenue Sharing program administered by the Treasury Department is the largest. Funds are allocated to States and local governments based on interdependent formulas in which the key variables are population, per capita income, and adjusted taxes.

In the areas of employment, there are two major programs -- the Comprehensive Employment and Training Act administered by Labor, and the vocational training program operated by the Office of Education in HEW. Both programs define eligibility and apportion assistance on the basis of poverty measures and unemployment rates. Other programs utilizing poverty measures as a basis for distributing aid are the Department of Agriculture's Food Stamp Program, the Community Mental Health Centers program of the National Institute of Mental Health, and HUD's Title I program under the Housing and Community Development Act of 1974. The latter program is designed primarily to improve housing for low to moderate income families in metropolitan areas.

All of these programs have one problem in common: significant changes have occurred since the 1970 Census, and as a result, there may be serious inequities in current patterns of assistance. Some of these changes, such as regional and metropolitan migration trends, are already well-documented, thanks to the Current Population Survey, but now the SIE offers reliable estimates of the net effect of these trends on the demographic and economic characteristics of individual States.

Will the SIE data supplant 1970 Census figures in existing allocation formulas? In some cases, such as Revenue Sharing, the answer is clearly no, since the SIE cannot begin to provide adequate estimates for the thousands of local governments involved. In other cases, programs may be wedded to 1970 poverty estimates because of required linkages with other types of data available only from the Decennial Census. This may apply to urban redevelopment programs, for example, since the SIE provides almost no information on housing characteristics.

As a general proposition, program managers are reluctant to make any change in grant procedures without a thorough study of the consequences, both statistical and political. With major tabulations of the SIE results now available, the statistical consequences are largely known, but more time may be required to assess the political ramifications. In this connection, it must be observed that delay serves to bring closer the time when the 1980 Census will make the question of using the SIE data entirely moot. Finally, two major evaluation studies bearing on the reliability of the SIE data and methods of developing sub-state poverty estimates are just now being completed.2/ These studies were mandated at the same time as the Survey, to guide the Congress in its deliberations on updating the Title I allocations. Clearly, the precedent set in this program area will carry considerable weight throughout the government.

Contributions to Policy Analysis

Before turning to an examination of some of the issues that are likely to shape the Title I decision, we wish to offer a few observations concerning the exceptional value of the SIE data for policy analysis.

Perhaps the first thing to be said is that the value of the SIE data to program planners and administrators is no accident: they played a major role in specifying the survey content. As a result, questions were added dealing with Food Stamp and public assistance recipiency, housing costs, liquid assets, child and adult disability, public and private health insurance coverage, and

changes in family composition affecting income reported for the previous year.

Broadly speaking, three types of policy analyses are being carried out: methodological studies, diagnostic studies, and simulation studies. Methodological studies have focused primarily on alternative measures of poverty. One of the richest areas in the SIE data for diagnostic studies is that of labor force participation, since extensive probes are utilized to develop a comprehensive picture of employment, job seeking, and reasons for periods of non-employment. Simulation studies utilizing SIE data have been devoted largely to evaluating welfare reform proposals. Working with data for individual households, and drawing on known relationships among various socio-economic variables such as age, occupation, and income, it is possible to simulate and hence "cost-out" or quantify the effects of different eligibility criteria and benefit levels. The ability to treat State of residence as a variable in these simulations has greatly improved their accuracy, since most existing welfare programs, including Aid for Dependent Children, Medicaid, and Food Stamps, are administered by the States, and benefit levels vary widely.

The SIE will continue for some years to be a prime source of data for policy analyses, but it is still reassuring to know that a successor survey is already in the works. Based on recommendations stemming from a comprehensive review of income statistics conducted by the Office of Management and Budget in the Spring of 1973, the Department of Health, Education, and Welfare is now planning a recurring Survey of Income and Program Participation, in conjunction with the Census Bureau. Many features of the SIE will be incorporated into the new survey, since the data, interview forms, sampling techniques, and field experience from the SIE have been consulted extensively in its development. The survey is slated to become operational in the 1980's.

Title I, ESEA

Although SIE data bear on policy issues in a number of different Federal programs, the survey's Congressional mandate was solely to update the poverty criterion specified in Title I of the Elementary and Secondary Education Act. Based on difficulties experienced earlier in converting from 1960 to 1970 Census estimates of poor families, the Congress decided that ten years is too long an interval between updates.

Title I participants are selected within school districts on the basis of various measures of educational need, but the estimated number of children in low income or "poverty" families is the key variable in the allocation of funds to States and local areas.3/ Thus, with the results of the SIE now in, it is possible to "cost out" the implications of changes since 1970 in the distribution of poverty children for Title I payments to the States. In this connection, it should be noted that Congress reserved for future deliberations the question of whether SIE estimates should supplant the 1970 figures in the Title I formula. These deliberations have now begun, and it is already clear that evaluations of the SIE results will figure conspicuously in the debate.

The SIE results show that there has been a significant shift in the distribution of children in poverty families between 1970 and 1976, based on income for the years 1969 and 1975 (Table 1). Comparatively fewer children now reside in the South (minus 12 percent), with the largest decreases occuring in Alabama (46 percent), Arkansas (26 percent), Louisiana (23 percent), and West Virginia (23 percent). Comparatively more poverty children reside in the industrial States. States showing the largest increases are New Jersey (35 percent), Illinois (39 percent), Michigan (23 percent), Pennsylvania (16 percent), and Ohio (16 percent). Among the twelve smallest States, seven show changes in excess of 30 percent, with large increases observed in Nevada, New Hampshire, and Vermont, and large decreases observed in Alaska, the District of Columbia, and the two Dakotas.

If sanctioned by Congress, these changes would have roughly proportional effects on Title I allotments. Based on FY 1977 figures, $131 million would have been re-allocated among States. While this is only about 8 percent of the total funding, the impact on individual States is considerable. In the two extreme cases Illinois would gain $27 million and Alabama would lose $22 million. Alabama would also experience the greatest proportional decrease (48 percent), while Nevada would receive the largest relative increase (54 percent).

Given the magnitude of these potential impacts, political considerations are likely to outweigh statistical ones in the final decision of the Congress. Nevertheless, statistical evaluations of the SIE results will figure in the debate, and in this connection there are three issues which are likely to receive close attention. These involve questions of sampling error, income reporting, and use of the SIE data for county-level estimates.

Despite the large size of the SIE sample -- over 150,000 households were interviewed -- the possibility of errors associated with sampling are likely to weigh heavily with Congress, particularly when translated into Title I allotment amounts. Since the sample was designed to minimize the relative error of the State estimates (the objective was to keep the coefficient of variation under 10 percent), the size of the absolute error in the larger States can be considerable. In the case of California, for example, one standard error in the estimate of children in poverty translates into $9.8 million of Title I funds, based on FY 1977 allotments. This amounts to more than two-thirds of the total cost of the survey. Thus, some will argue that there is a serious disproportion between the accuracy of the SIE (and its associated costs) and the amounts at risk under the Title I program.4/ Given the logic of the SIE sample design, there are a great many statements which can be made concerning the likelihood of error, some of which will doubtless excite concern in the Congress. Thus, for ex-

TABLE 1: CHANGES IN RELATIVE SHARES OF POVERTY CHILDREN AND TITLE I FUNDS FOR STATES AND REGIONS: 1970 TO 1976

REGION State	Percent of total poverty children in United States			Title I allocations for FY 1977 (in millions)		Difference	
	1970 Census	SIE (1976)	Percent increase	Actual (based on '70 Census)	Hypothetical (using SIE)	Amount	Percent
United States, Total	100.00	100.00	--	$ 1,653	$ 1,653	--	--
NORTHEAST	16.21	18.64	15.0	380	417	37	+10
Maine	0.47	0.54	14.9	6	7	1	+16
New Hampshire	0.19	0.28	47.4	3	4	1	+32
Vermont	0.17	0.28	64.7	3	5	2	+48
Massachusetts	1.52	1.73	13.8	32	35	3	+9
Rhode Island	0.32	0.30	- 6.3	7	6	-1	-12
Connecticut	0.72	0.83	15.3	15	18	3	+20
New York	6.84	7.35	7.5	184	189	5	+3
New Jersey	2.02	2.72	34.7	46	60	14	+32
Pennsylvania	3.96	4.61	16.4	84	93	9	+11
NORTH CENTRAL	20.00	22.12	10.6	363	400	37	+10
Ohio	3.55	4.12	16.1	52	58	6	+11
Indiana	1.60	1.69	5.6	21	23	2	+9
Illinois	3.93	5.46	38.9	88	115	27	+31
Michigan	2.86	3.53	23.4	70	84	14	+19
Wisconsin	1.35	1.49	10.4	27	29	2	+6
Minnesota	1.28	1.22	- 4.7	25	25	--	-2
Iowa	0.94	0.75	-20.2	15	12	-3	-20
Missouri	2.25	2.20	- 2.2	31	29	-2	-4
North Dakota	0.36	0.25	-30.6	5	3	-2	-34
South Dakota	0.44	0.30	-31.8	5	4	-1	-24
Nebraska	0.60	0.51	-15.0	10	8	-2	-24
Kansas	0.84	0.60	-28.6	14	10	-4	-34
SOUTH	49.55	43.43	-12.4	660	571	-89	-13
Delaware	0.23	0.20	-13.0	5	4	-1	-18
Maryland	1.52	1.46	- 3.9	28	28	--	--
District of Columbia	0.48	0.32	-33.3	10	7	-3	-31
Virginia	2.78	2.18	-21.6	39	31	-8	-21
West Virginia	1.38	1.07	-22.5	18	12	-5	-27
North Carolina	4.06	3.10	-23.6	50	39	-11	-22
South Carolina	2.69	2.32	-13.8	34	29	-6	-17
Georgia	3.82	3.57	- 6.5	50	44	-6	-11
Florida	3.89	5.36	37.8	61	80	19	+31
Kentucky	2.71	2.39	-14.8	34	30	-4	-12
Tennessee	3.18	2.74	-13.8	41	34	-7	-16
Alabama	3.53	1.91	-45.9	46	24	-22	-48
Mississippi	3.40	2.72	-20.0	42	34	-8	-20
Arkansas	2.01	1.49	-25.9	26	18	-8	-29
Louisiana	4.01	3.08	-23.3	53	39	-14	-27
Oklahoma	1.59	1.22	-23.3	19	15	-4	-21
Texas	8.27	8.30	0.3	104	103	-1	-2
WEST	14.26	15.80	10.8	248	265	17	+7
Montana	0.32	0.32	0.0	5	5	--	-4
Idaho	0.31	0.32	3.2	4	4	--	+4
Wyoming	0.13	0.11	-15.4	2	2	--	-4
Colorado	0.93	0.90	- 3.2	16	14	-2	-11
New Mexico	1.05	1.09	3.8	14	14	--	-4
Arizona	1.09	1.30	19.3	15	18	3	+17
Utah	0.40	0.35	-12.5	6	5	-1	-13
Nevada	0.14	0.22	57.1	2	3	1	+54
Washington	1.04	1.14	9.6	19	21	2	+11
Oregon	0.70	0.60	-14.3	15	13	-2	-17
California	7.74	9.09	17.4	141	158	17	+12
Alaska	0.16	0.09	-43.7	3	2	-1	-23
Hawaii	0.25	0.27	8.0	6	6	--	--

ample, the probability of ten or more State estimates being off by more than 10 percent is .99, and conversely, the chance of estimates for all 50 States and the District of Columbia being within 10 percent is one in a billion.

It is interesting to note that while the SIE sample design equitably distributes the risk of error across States, it is not efficient from the standpoint of targeting Title I dollars. To minimize the number of Title I dollars misdirected as a result of sampling error, the sample size would have to have been proportional to the estimated number of children in poverty as well as the inverse of the poverty rate.5/

Based on reinterview studies, it appears that income reporting in the SIE was relatively more complete than that generally obtained in the Decennial Census or the Current Population Survey. Since the effect of unreported income is to inflate estimates of poverty, this means that the SIE provides the best estimate of the total number of children age 5-17 in families falling below the poverty line, but it also means that SIE-Census comparisons provide a distorted picture of changes over time. Using the results of the March CPS for 1970 and 1976 as a bridge between the 1970 Census and the SIE, we estimate that if the rate of nonreporting of income in the SIE had been comparable to that in the 1970 Census, the number of children in poverty would have been 23 percent higher. This means that instead of the observed decrease of 7.4 percent since 1970, we might have obtained an increase of 14.3 percent (Table 2), a shift of nearly 22 percentage points. While these calculations are admittedly speculative, and depend on the assumption that no change in income reporting has occurred between 1970 and 1976 in the CPS, the potential magnitude of these shifts is enough to justify serious concern. The basic intent of the SIE was to obtain estimates which would help to bridge the gap between the 1970 and 1980 Censuses, but on this evidence, it seems likely that the 1980 Census will show substantially greater numbers of children in poverty than could have been expected from the SIE results.

One final point should be noted concerning the question of sampling error. Even when the 1970 Census figures for children in poverty are rateably reduced to yield the same total as the SIE, most of the major changes observed at the State level are highly significant. Thus, for Illinois and New Jersey -- the two States which stand to gain the largest amounts -- the observed differences are respectively 4.0 and 3.3 times the standard error of the estimates. This also holds true for States which would experience the largest relative increases: Vermont, Nevada, and New Hampshire all exhibit differences in excess of 3.5 standard errors.

As we have indicated, there is some question of whether the SIE estimates are sufficiently accurate at the State level, and the Title I allocation process requires estimates of poverty children down to the county level. Thus, if Congress were to authorize the use of the SIE estimates in determining the amount each State receives, the problem of how the States would suballocate to the county level still remains. One possible solution, explored by Abduhl Kahn and Herman Miller in connection with a special study mandated by Congress, is to develop synthetic estimates. This method applies trend data for metropolitan and non-metropolitan counties at the State level as an adjustment to 1970 data for individual counties. Two serious limitations of this approach are: 1) measures of reliability are not calculable for such estimates, and 2) the method may be perceived as open to manipulations designed to produce preconceived results. Logically, the easy way out would be to let the States work out their own methods of allocating funds down to the county level, but Congress is very reluctant to do this, for fear that some States would re-direct Federal funds away from low-income areas.

TABLE 2: ADJUSTMENTS IN SIE ESTIMATE OF POVERTY CHILDREN FOR COMPARABILITY WITH 1970 CENSUS, BASED ON COMPARISONS WITH CURRENT POPULATION SURVEY ESTIMATES.

	Related children 5-17 in poverty (thousands)		
	Actual estimate	Adjusted estimate*	Percent change
1970 Census.....	7,700	7,700	--
1970 March CPS....	7,000	7,700	+10.0
1976 March CPS....	8,000	8,800	+10.0
1976 SIE.....	7,132	8,800	+23.4
Percent change '70 Census to SIE.....	-7.4	+14.3	

* Estimates are adjusted for comparability with the 1970 Census. CPS income reporting is assumed to have remained the same between 1970 and 1976; thus, our calculations suggest that the 1970 Census methodology would have yielded an estimate 10 percent higher than the CPS in 1976 just as it did in 1970.

Conclusion

Data from the SIE will be used over the next four to five years by policy analysts in a number of different program areas within the Department of Health, Education, and Welfare, including bilingual education, education of handicapped children, welfare reform, and postsecondary education. As mentioned earlier, at least four other departments -- Labor, Treasury, Agriculture, and Housing and Urban Development -- also plan to make special use of the SIE data. Clearly, then, there is no question about the benefits of the survey exceeding its cost. Relative to the annual appropriations of the programs benefiting, the $14 million cost of the SIE is a nearly invisible fraction. Relative to the original purpose of updating the Title I allocations, however, these are "fringe" benefits, based largely on add-ons to the scope of the survey.6/ Thus, while the

Federal government can congratulate itself on having successfully exploited the opportunity afforded by this mandated survey, the possibility that it may never be used for the purpose originally intended must give pause for reflection.

In retrospect, it was a mistake for the Congress to defer a decision on the use of the SIE. As a result, the question of the required degree of accuracy of the SIE estimates was not conclusively resolved, and now the need for further deliberations means that Title I allocations cannot be updated until FY 1980 at the earliest -- just two years from the time when the results of the 1980 Census will become available.

At the present time, it appears that the SIE data will be used to up-date State allocations, but reaching agreement on this is likely to require a hold-harmless provision plus an increase in Title I funding. Based on the funding level in FY 1977, a full hold-harmless would cost an additional $131 million on a base of $1.6 billion. Within-State allocations will probably continue to be determined on the basis of the 1970 data for counties.

Looking to the future, there is a serious question of whether the 1980 Census will produce poverty estimates comparable to those of the SIE. If past experience is any guide, under-reporting of income is likely to inflate the Census estimates of poverty. There is even some question as to whether the 1980 income data will be comparable to those from 1970, since the Census Bureau is experimenting with a simplified income question for use on the complete-count form, and plans to ask much more detailed income questions in a follow-on survey.

With the authorization of a mid-decade or quinquennial census, major Federal programs will no longer have to endure a ten-year hiatus between reliable measures of the social and economic conditions they are designed to ameliorate. It is likely, however, that special surveys will continue to be required in order to provide the necessary detail for narrow-gauge programs targeted on particular types of disadvantaged groups. In this connection, we believe the SIE will serve as a useful model of the benefits to be realized by pooling needs and sharing costs.

FOOTNOTES

1/ Section 822a of the Education Amendments of 1974 -- Public Law 93-380.

2/ These reports, now being prepared at HEW and the Census Bureau, are entitled (1) "The Survey of Income and Education" and (2) "Counting Poor School Children".

3/ Slightly simplified, the allocation formula may be described as "eligibles" times "payment rate" times "rateable reduction", where: (1) eligibles are the sum of the children age 5-17 in poverty families as defined in the 1970 Census plus two-thirds of children in families above the poverty line line receiving AFDC payments plus children in foster homes or institutions for the neglected and delinquent; (2) the payment rate is 40 percent of each State's current educational expenditure per pupil but not less than 80 and not more than 120 percent of national average expenditures, and (3) the rateable reduction is the ratio of the current appropriation to the amounts otherwise authorized.

4/ Achieving a coefficient of variation of 2½ percent was estimated to cost between $50 and $100 million. This was judged to be excessive, in part because added costs would have come out of Title I program money, and the program is already funded at substantially below the estimated level of need.

5/ To minimize the variance of the individual State estimates, the fraction of the total sample (n) allocated to a given State (s) with C_s estimated children in poverty constituting Ps proportion of all children is given by the proportion:

$$\frac{C_s}{C_n} \sqrt{\frac{P_s \quad (1-P_s)}{P_n \quad (1-P_n)}}$$

6/ In the case of data needed for the bilingual education program, it was necessary not only to add questions dealing with limitations in the use of English, but also to expand the SIE sample in selected States in order to obtain estimates of sufficient reliability for children of limited English-speaking ability.